SPR Day 1

Goals: Why take this class?

What are the fundamental rules of probability?

Concepts: discrete random variables
sample space
probability mass function
joint distribution
conditional distribution
marginal distribution
sum rule
product rule
Bayes rule
independence
conditional independence

Probability Theory

Hopefully, mostly a review/refresher

See Fig 1.9 in Bishop PRML

@ = apple

o = orange

Consider the following stochastic procedure

1) select a random bucket, either red or blue
   choose red 40% of the time
   blue 60%
2) from that bucket, select one fruit uniformly at random
3) report that fruit as "apple" or "orange", then replace it

What is the probability of picking an apple?

Approach

(0) Break problem down into easier parts

(1) Clearly define random variable(s) including sample space and distribution

(2) Apply necessary laws of probability

Events of interest

random variable, value

Which box is picked? $B = \text{red}$ with prob. 40%

or

$B = \text{blue}$ 60%

sample space $= \{\text{red}, \text{blue}\}$
is the set of all possible outcomes

Which fruit is picked in red box?

$F = o$ with prob. $2/8$

$F = @$ with prob. $6/8$

Which fruit is picked in blue box

$F = o$ w/ prob. $3/4$

$F = @$ $1/4$

In General

Sample space defines all possible outcome values for a random variable.

Probability of each outcome must be between 0 & 1

set of (unordered) possible outcomes that is
(1) mutually exclusive: each outcome is distinct/non-overlapping
(2) collectively exhaustive:
all possible outcomes represented

If Ω is sample space for random variable X, then,

$$\sum_{x \in \Omega} Pr(X = x) = 1$$

Joint Probability

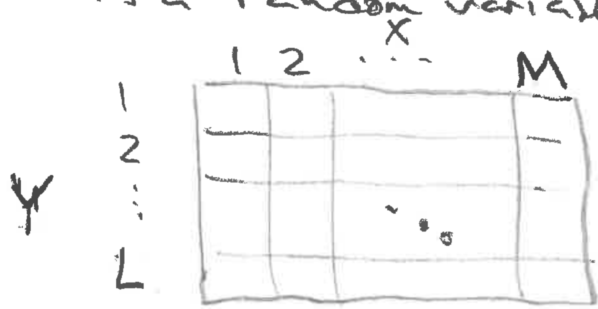
Setup:

X is random variable w/ sample space
$$\mathcal{X} = \{x_1, x_2, \cdots x_M\}$$

Y is random variable w/ sample space
$$\mathcal{Y} = \{y_1, \cdots y_L\}$$

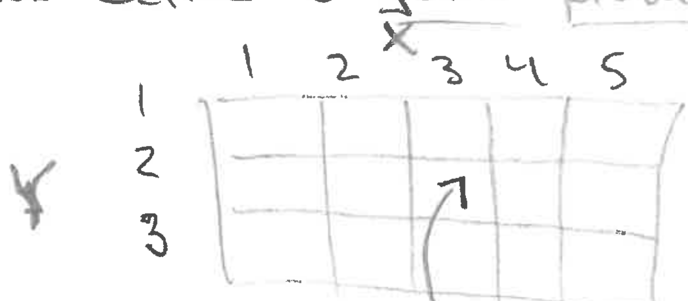
Now consider the unordered **pair** of X and Y together.

This pair is a random variable w/ sample space



with $M \times L$ elements

We can thus define a joint probability table



where the sum of all entries must be 1

we write an entry as $P_r(X = x_2, Y = y_3)$

read as "and"

Conditional Probability

We want to ask a question like this:

given that I picked the red box,
what is the probability that I'll
pick out an apple?

Write as

$$\Pr(F = @ \mid B = \text{red})$$

read as

"probability of the event [fruit is an apple], given that the event [box is red] is true"

Two fundamental rules

allow us to relate formally between
- joint events
- conditional events
- marginal events

Two Rules:

Sum Rule

$$\Pr(X = x_i) = \sum_{y \in \mathcal{Y}} \Pr(X = x_i, Y = y)$$

Sometimes called "marginal" probability,
act of summing away Y is called "marginalizing"
imagine writing the sum's result in the "margin" of the joint table

Product Rule

$$\Pr(Y = y_j \mid X = x_i) \Pr(X = x_i) = \Pr(X = x_i, Y = y_j)$$

prob of picking this row, given this column

prob of this column

Can check by thinking of joint probability table as normalized counts (see Bishop Eq 1.9)

works for many vars: $P(X_1, \cdots X_N) = \prod_{n=1}^{N} p(X_n \mid X_1 \cdots X_{n-1})$ chain rule

Bayes Rule

can be derived from the product rule

$$\Pr(Y=y \mid X=x) = \frac{\Pr(X=x \mid Y=y)\Pr(Y=y)}{\Pr(X=x)}$$

why? multiply both sides by $\Pr(X=x)$

$$\Pr(Y=y \mid X=x)\Pr(X=x) = \Pr(X=x \mid Y=y)\Pr(Y=y)$$

$$\Pr(Y=y, X=x) = \Pr(X=x, Y=y)$$

true by symmetry

Exercise

1) Compute the following from the joint distribution

$\Pr(Y=y)$

$\Pr(X=x)$

2) Now Compute each of these from the joint (can reuse result from #1)

$\Pr(Y=y \mid X=x)$

Return to fruit baskets

$Pr(B = \text{red}) = \frac{4}{10}$

$Pr(B = \text{blue}) = \frac{6}{10}$

$Pr(F = o \mid B = \text{red}) = 3/4$

$Pr(F = @ \mid B = \text{red}) = 1/4$

$Pr(F = o \mid B = \text{blue}) = 1/4$

$Pr(F = @ \mid B = \text{blue}) = 3/4$

Exercise

Use sum & product rules to give overall prob. of choosing an apple

$$Pr(F = @) = ?$$

Should be $11/20$
see (1.22)

Use sum & product rules to compute

$$Pr(B = \text{red} \mid F = o) = ?$$

Should be $2/3$
see (1.23)

How did we have enough information to solve these?
our model defined a joint distribution

Independence & Conditional Independence

Two random variables X and Y are "independent" if the distribution of Y given knowledge of X is the same as the marginal distribution of Y

$$\Pr(Y=y \mid X=x) = \Pr(Y=y) \quad \text{for all } y, x$$

That is, the distribution of Y never depends on the value of X

The joint distribution of two indep. r.v.s looks like a product of marginals

$$\Pr(X=x, Y=y) = \Pr(Y=y \mid X=x)\Pr(X=x) \quad \text{product rule}$$

$$= \Pr(Y=y)\Pr(X=x) \quad \text{defn of Independence}$$

Conditional Independence

We say two rand vars X and Y are "conditionally independent" given a third random variable Z iff

$$\Pr(Y=y \mid X=x, Z=z) = \Pr(Y=y \mid Z=z) \quad \text{for all } x, y, z \text{ values}$$

again, this means the joint distr. looks like

$$\Pr(X=x, Y=y, Z=z) = \Pr(Y=y \mid Z=z)\Pr(X=x \mid Z=z)\Pr(Z=z)$$

Expectation

Given a discrete random variable X,
we might have a function of X in mind for a task
(e.g. the cost of the fruit selected,
the points I get for each dice role,
etc)

Let $f(x)$ map from sample space of X
to a real value or
vector of real values.

We want to know the "average" value of $f(X)$
weighting each outcome by its probability.

$$\mathbb{E}[f(X)] = \sum_{x \in \mathcal{X}} p(X=x) f(x)$$

When each value x is represented as a real value
or vector,
we can compute the mean of random variable X

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} p(X=x)\, x$$